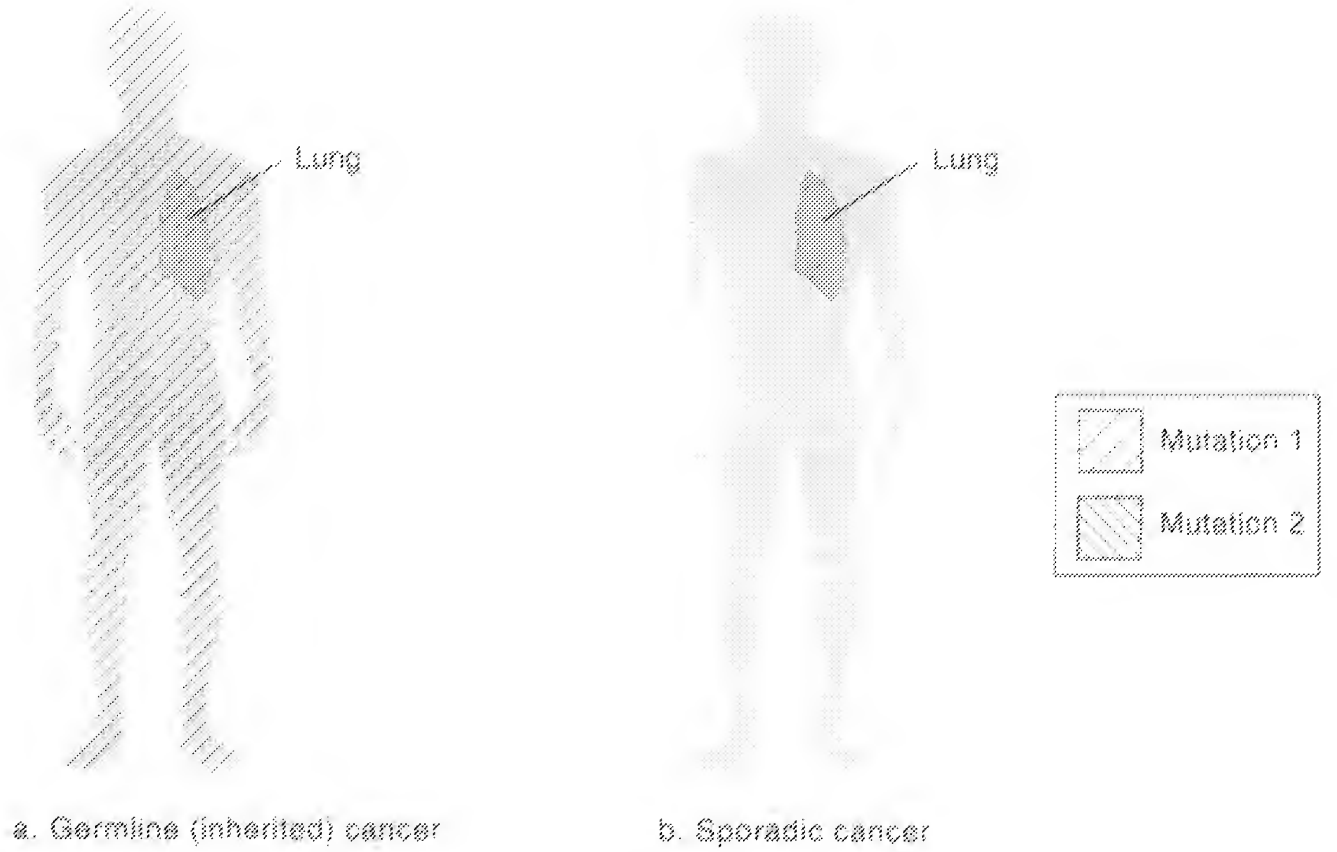
Universal DNA Arrays for Improved Cancer Care

Dr. Francis Barany
Director: Hidden Cancer Project
Department of Microbiology & Immunology
Weill Cornell Medical College
1300 York Avenue
New York 10065
barany@med.cornell.edu

Cancer will affect 1 in 3

- 25% of cancers run in families (inherited).
- 75% of cancers are sporadic, i.e. anyone can get it.
- Every year, 1.3 million Americans will be diagnosed with cancer, and 550,000 will die from cancer.

Germline vs. Sporadic Cancer



a. Germline (inherited) cancer

b. Sporadic cancer

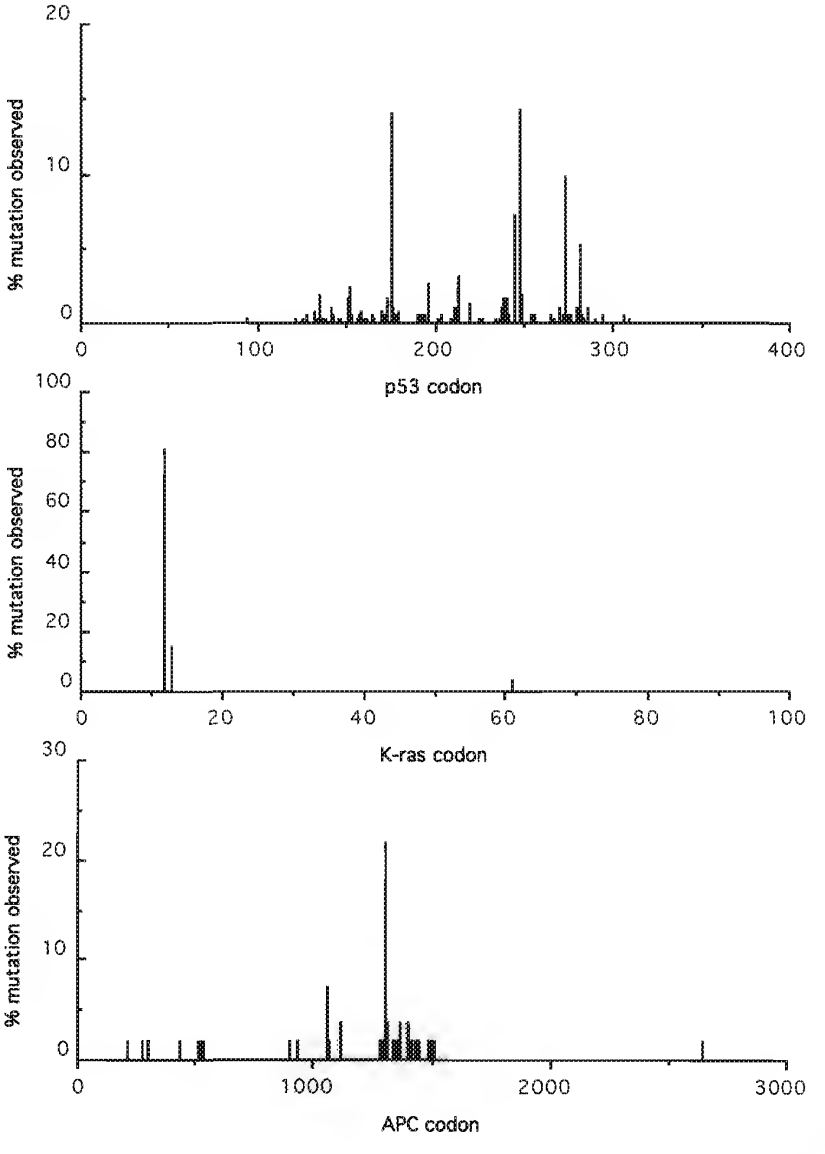
Mutations observed in Colon Tumors

Challenges:

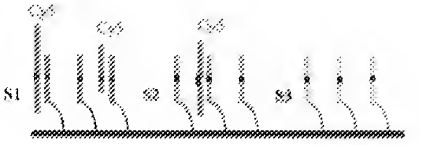
1. Stromal contamination: Mutation may be present in only 15%-25% of DNA.

2. Multiple, and closely spaced mutations in multiple genes.

3. For early detection, need to find one mutant in 100 wild-type.

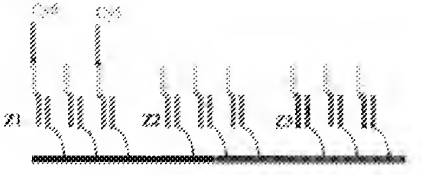


Comparison of Arrays for Identifying Mutations

Hybridization Arrays: 

• Discrimination based on match vs. mismatch hybridization; probes are very similar in sequence.

• Tm (melting temperature) of probes depend on target; considerable variation.

Universal Arrays: 

• Discrimination based on match vs. mismatch ligation.

• Capture probes designed to have distinct sequences with no cross-hybridization.

Identifying Mutations by Hybridization: 1

```
Target: 5'-CATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGA
Probes:              3'-TTTATAXTAGAAACC (SEQ. ID NO:9)
                      3'-TTATAGXAGAAACCA (SEQ. ID NO:10)
                       3'-TATAGTXGAAACCAC (SEQ. ID NO:11)
                        3'-ATAGTAXAAACCACA (SEQ. ID NO:12)
                         3'-TAGTAGXAACCACAA (SEQ. ID NO:13)
                          3'-AGTAGAXACCACAAA (SEQ. ID NO:14)
                           3'-GTAGAAXCCACAAAG (SEQ. ID NO:15)
                            3'-TAGAAAXCACAAAGG (SEQ. ID NO:16)
                             3'-AGAAACXACAAAGGA (SEQ. ID NO:17)
```

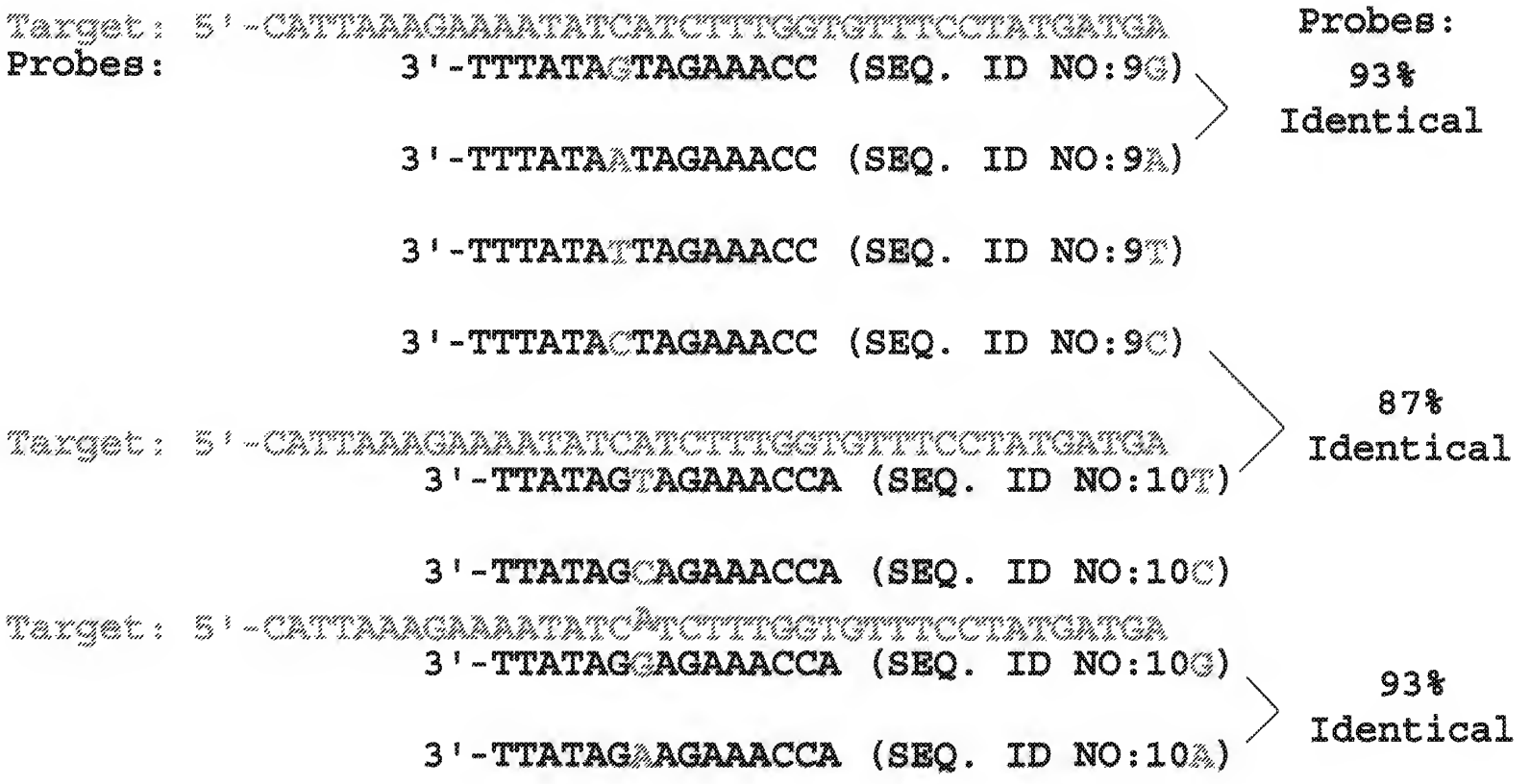
Identifying Mutations by Hybridization: 2

```
Target:  5'-CATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGA
Probes:              3'-TTTATAGTAGAAACC (SEQ. ID NO:9G)
                     3'-TTTATAATAGAAACC (SEQ. ID NO:9A)
                     3'-TTTATATTAGAAACC (SEQ. ID NO:9T)
                     3'-TTTATACTAGAAACC (SEQ. ID NO:9C)

                      3'-TTATAGTAGAAACCA (SEQ. ID NO:10T)
                      3'-TTATAGCAGAAACCA (SEQ. ID NO:10C)
                      3'-TTATAGGAGAAACCA (SEQ. ID NO:10G)
                      3'-TTATAGAAGAAACCA (SEQ. ID NO:10A)

                       3'-TATAGTAGAAACCAC (SEQ. ID NO:11A)
                       3'-TATAGTTGAAACCAC (SEQ. ID NO:11T)
                       3'-TATAGTCGAAACCAC (SEQ. ID NO:11C)
                       3'-TATAGTGGAAACCAC (SEQ. ID NO:11G)
```

Identifying Mutations by Hybridization: 3

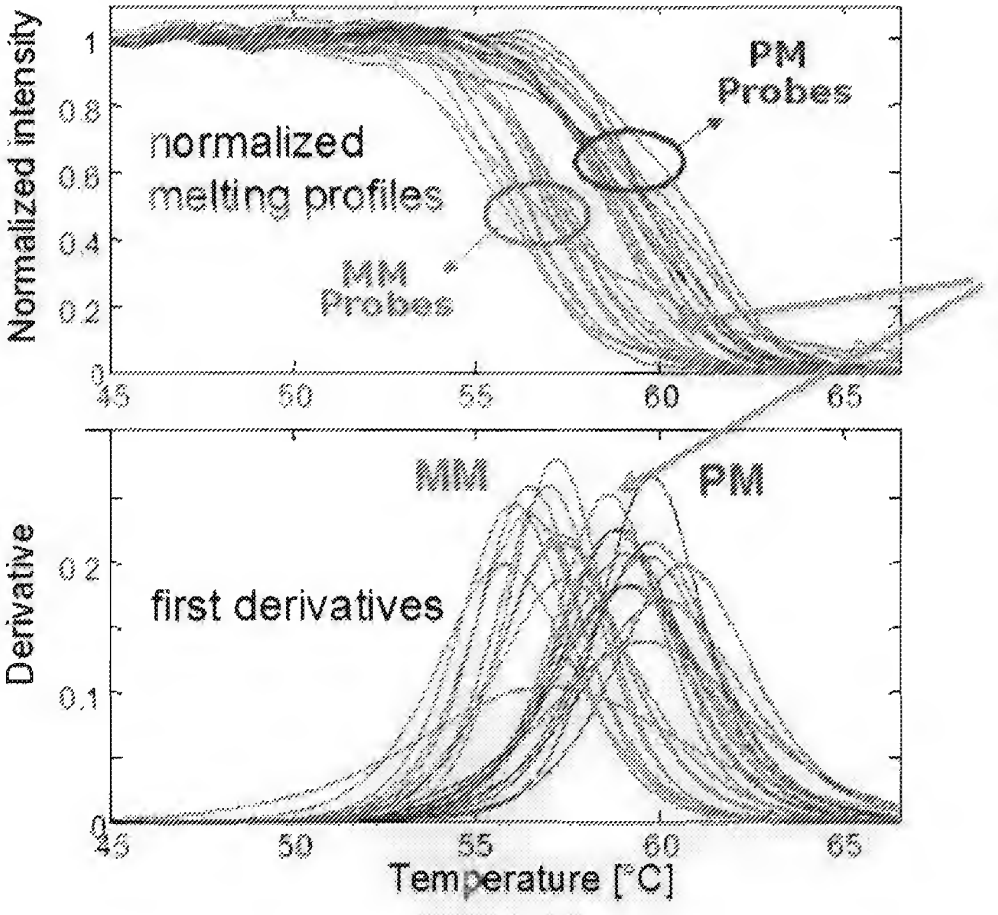


Oligonucleotide microarrays: widely applied--poorly understood.

Microarray technology, which has been around for almost two decades, now provides an indispensable service to the biomedical research community. Soaring demand for high-throughput screening of genes potentially associated with cancer and other diseases…, have substantially opened up the application of this technology to many fields of science. Yet, despite this significant progress, the fundamental understanding of the pillars of this technology, have been largely unexplored, in particular for oligonucleotide-based microarrays. In fact, most of the current approaches for the design of microarrays are based on 'common-sense' parameters, such as guanine- cytosine content, secondary structure, melting temperature or possibility of minimizing the effects of nonspecific hybridization. However, recent experiments suggest that these are inadequate. Here we discuss these results, which challenge the basic principles and assumptions of oligonucleotide microarray technology. It is clear that more systematic physicochemical studies will be required to better understand the hybridization and dissociation behavior of oligonucleotides.

Pozhitkov AE, Tautz D, Noble PA.
Brief Funct Genomic Proteomic. 2007 Jun;6(2):141-8.

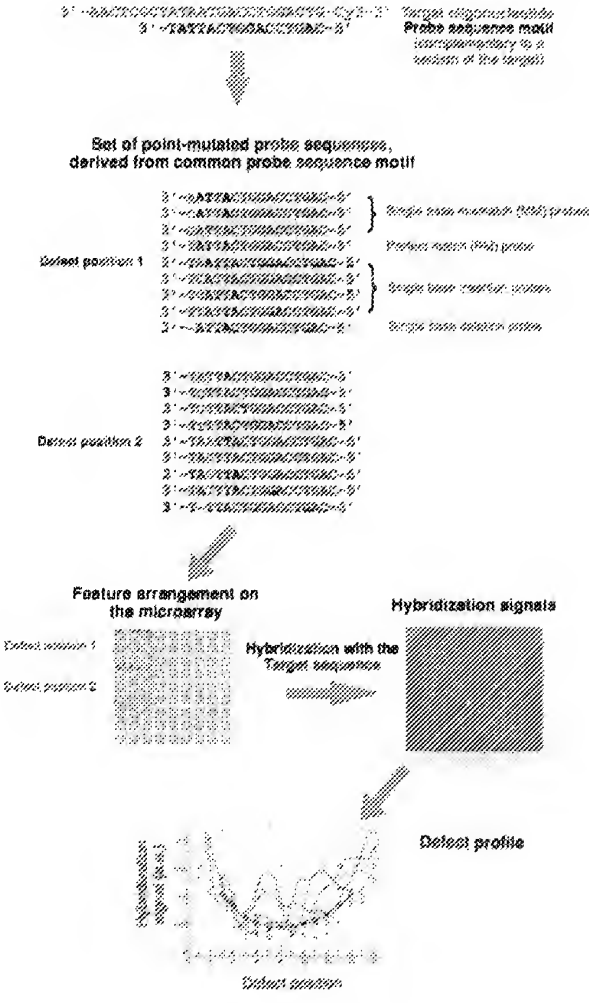
"Normalized" Melting Profiles of oligonucleotides in solution



Even in solution, some mismatched probes (MM) have melting curves like perfect matched (PM) probes.

http://www.genewave.com/images/manager/hyblive_schema1.JPG

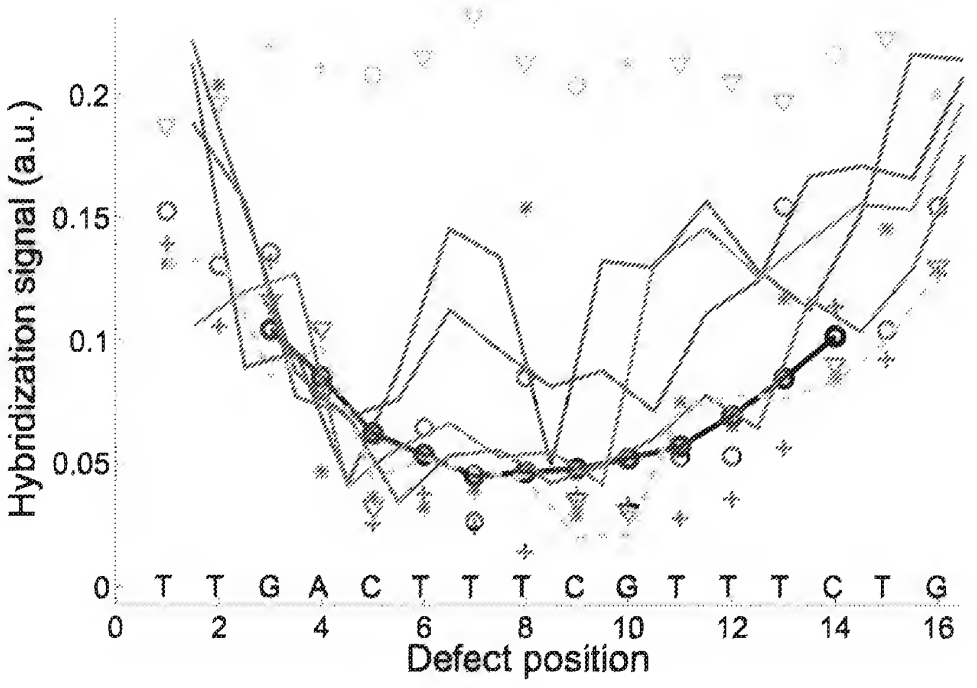
Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges.



Design of the experiment. A comprehensive set of point-mutated probes is derived from a common probe sequence motif which is complementary to the target sequence. Probe sequences are shown for the first two defect positions only. To enhance quantitative analysis probe sequences are arranged on the microarray as a compact feature block. Hybridization signals from hybridization with the target sequence are plotted versus defect position. The defect profile shows relative hybridization affinities depending on the probe sequence motif, defect type and defect position.

Naiser T, Ehler O, Kayser J, Mai T, Michel W, Ott A.
BMC Biotechnol. 2008 May 13;8:48

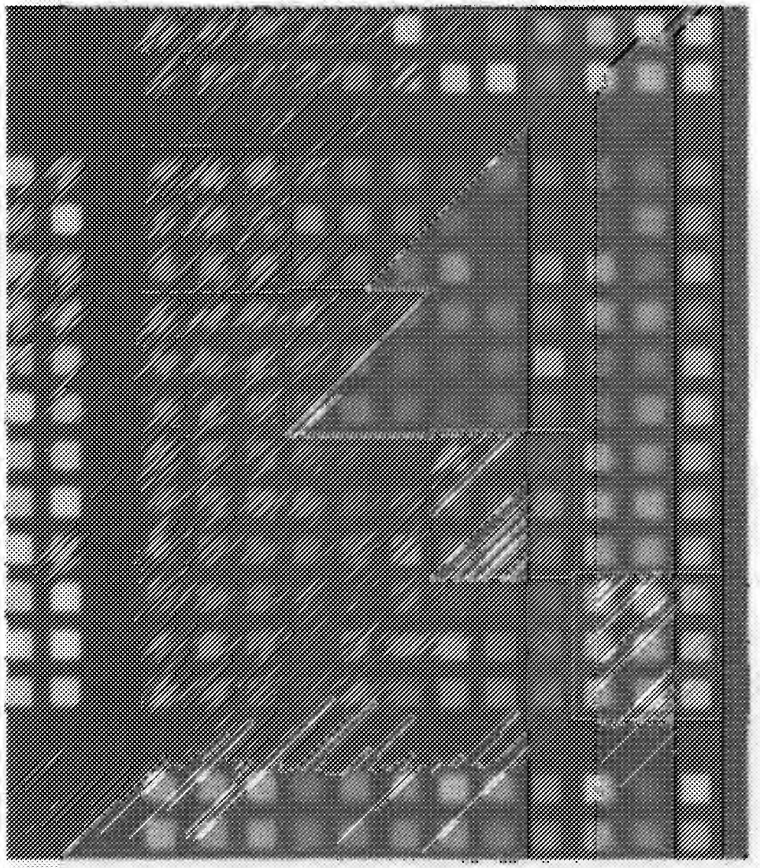
Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges.



Direct comparison of single base mismatches, insertions and deletions. The 16 mer probe sequence motif 3'- TTGACTTTCGTTTCTG-5' is complementary to the target BEI. Hybridization signals (data processing: raw fluorescence intensities; solution-background correction) of single base mismatch probes with substituent bases A (red crosses), C (green circles), G (blue stars), T (cyan triangles), running average of mismatch intensities (black line); perfect match probe signals (grey symbols) single base insertion probes (solid lines) with insertion bases A (red), C (green), G (blue), T (cyan). Hybridization signals of single base deletions (orange dashed line) are comparable to that of mismatches at the same position. Increased hybridization signals of certain insertion defects are due to positional degeneracy of base bulges (see discussion).

Naiser T, Ehler O, Kayser J, Mai T, Michel W, Ott A. BMC Biotechnol. 2008 May 13;8:48

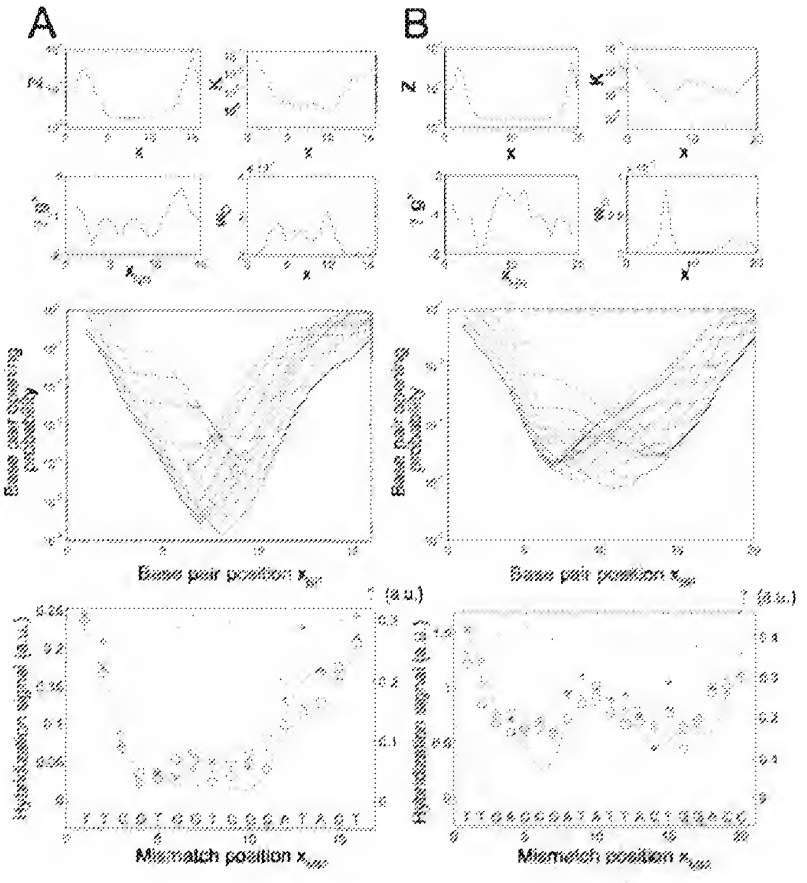
Position dependent mismatch discrimination on DNA microarrays - experiments and model.



Fluorescence micrograph (taken with an Olympus IX81 epi-fluorescence microscope and a Hamamatsu EM- CCD camera) of a microarray feature-block comprising variations of the 16 mer probe sequence motif 3'-TATTACTGGACCTGAC-5'. Microarray hybridization was performed with the 5'-Cy3-labeled RNA oligonucleotide target 3'-AACUCGCUAUAAUGACCUGGACUG-5' (target concentration: 1 nM in 5 × SSPE, pH 7.4, 0.01% Tween-20, T = 30°C). Each 3 × 3 sub-array comprises (randomly arranged) one perfect matching probe, three single base mismatch probes, four insertion probes and one single base deletion probe. In Fig. 2A the hybridization signals (fluorescence intensities, averaged over the center of the microarray features) are plotted versus the defect position. The size of each microarray feature is 21 μm and the pitch of the array is 35 μm. The significantly brighter feature-block at left comprises variations of the 20 mer probe sequence motif 3'-TTGAGCGATATTACTGGACC-5'.

Naiser T, Kayser J, Mai T, Michel W, Ott A.
BMC Bioinformatics. 2008 Dec 1;9:509.

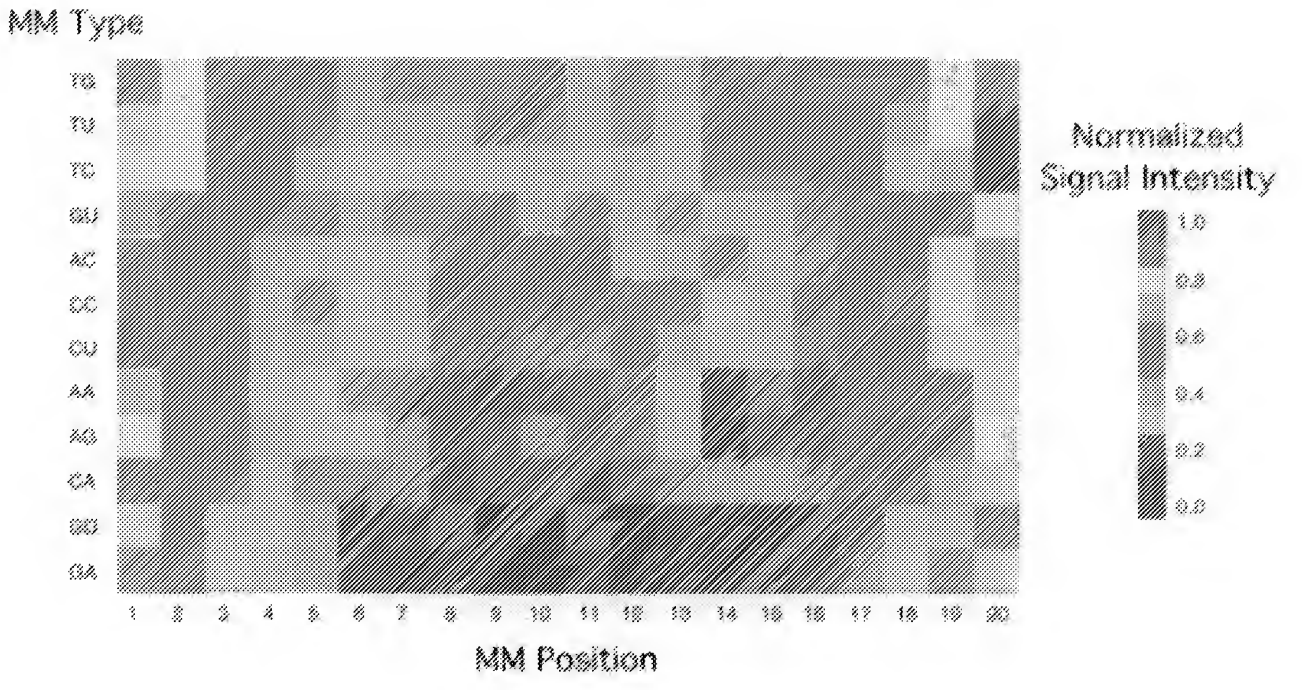
Position dependent mismatch discrimination on DNA microarrays - experiments and model.



Comparison of simulation results with the experimentally determined hybridization affinities for two probe sequence motifs (A) and (B). The four small sub-figures in the top section (from top left to bottom right) show the partition function Z and the duplex binding constant K as a function of defect position x (semi-logarithmic plots), the NN-free energies Δg° of particular NN-pairs as a function of NN-pair position xNN, and the statistical weight for complete duplex dissociation wD as a function of defect position. Irregularities in Z(x) at the duplex ends are an artifact caused by the fact that only a single NN-pair is affected by a MM-base pair at the duplex end. The middle sub-figure shows the base pair opening probabilities (the fraction of strands in which the corresponding base pair at position xBP is unzipped) as a function of the defect position. The spectrum of differently colored curves encodes the different defect positions xMM (red – defect at left end; purple – defect at right duplex end). The bottom sub-figure compares the experimentally determined MM defect profile (mismatched base: A (red cross), C (green circle), G (blue star), T (cyan triangle); gray symbols correspond to PM probes) with the simulated MM defect profile θ(x) (dashed orange line). With Δgdef = 1 kcal/mol (at the simulation temperature of 325 K) and an error rate of 12 percent (per synthesis step) the calculated defect profile θ(x) matches well the experimental data.

Naiser T, Kayser J, Mai T, Michel W, Ott A.
BMC Bioinformatics. 2008 Dec 1;9:509.

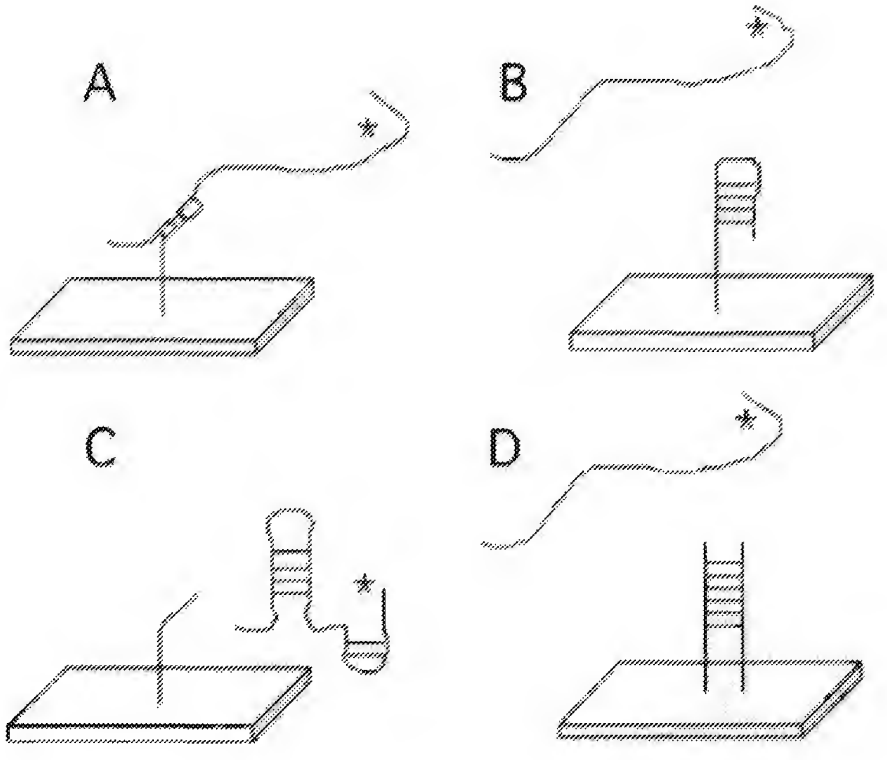
Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted.



Heat map of MM type by position as a function of average signal intensity, normalized to the signal intensity of the PM duplex. Each box represents at least 120 replicates.

Pozhitkov A, Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M, Tautz D. Nucleic Acids Res. 2006 May 17;34(9):e66.

Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted.

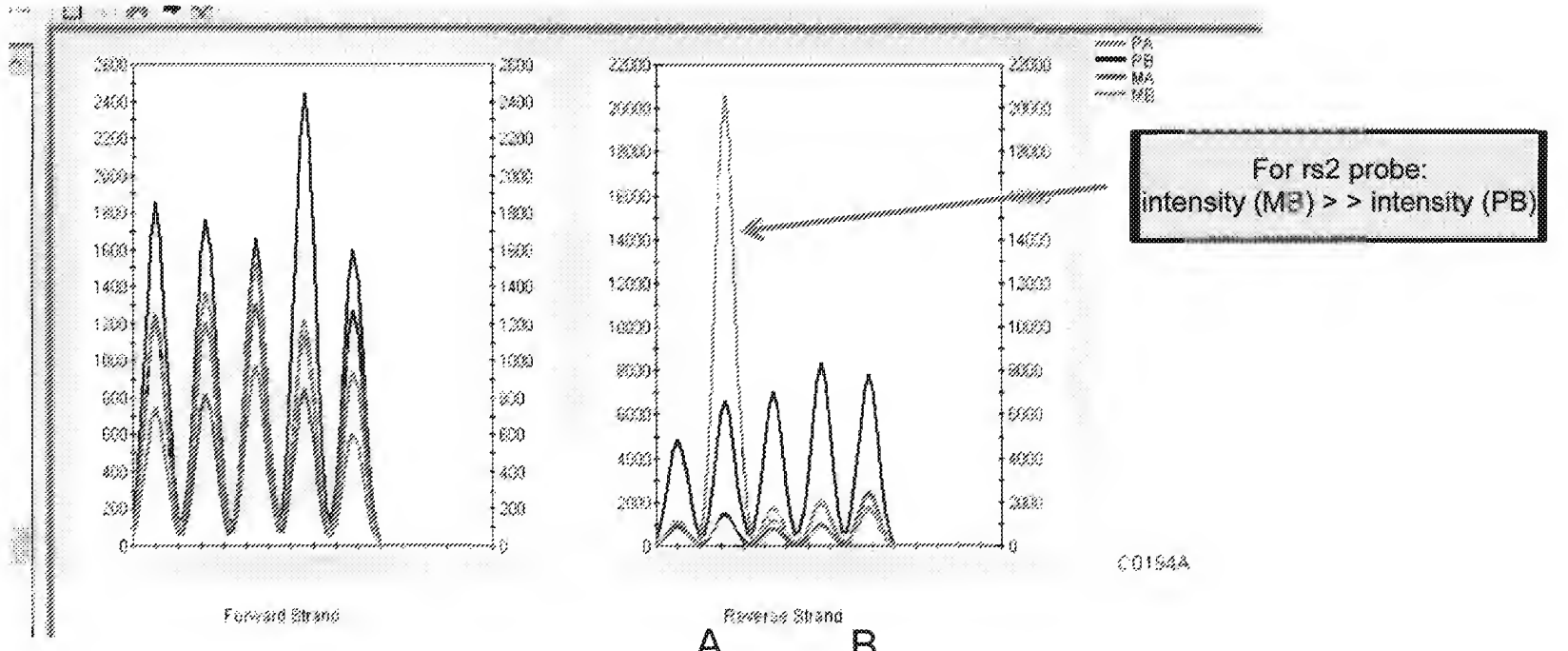


Depiction of four competitive processes on signal intensity values. Each panel shows a labeled (*) target and an immobilized probe on a micro- array. (A) hybridization of a target to a probe; (B) probe self-folding; (C) folding of the target and (D) dimerization of adjacent probes.

Pozhitkov A, Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M, Tautz D. Nucleic Acids Res. 2006 May 17;34(9):e66.

Xba 240 SNP array analysis of CRC sample C0194A

▶ Testing the genotype for SNP_A-1713319



▶ Probesets designed for:

tatgttacatcagttactcctttca[**A/G**]tatagattaggtttttaagtcctcc

25-mer tiled probes

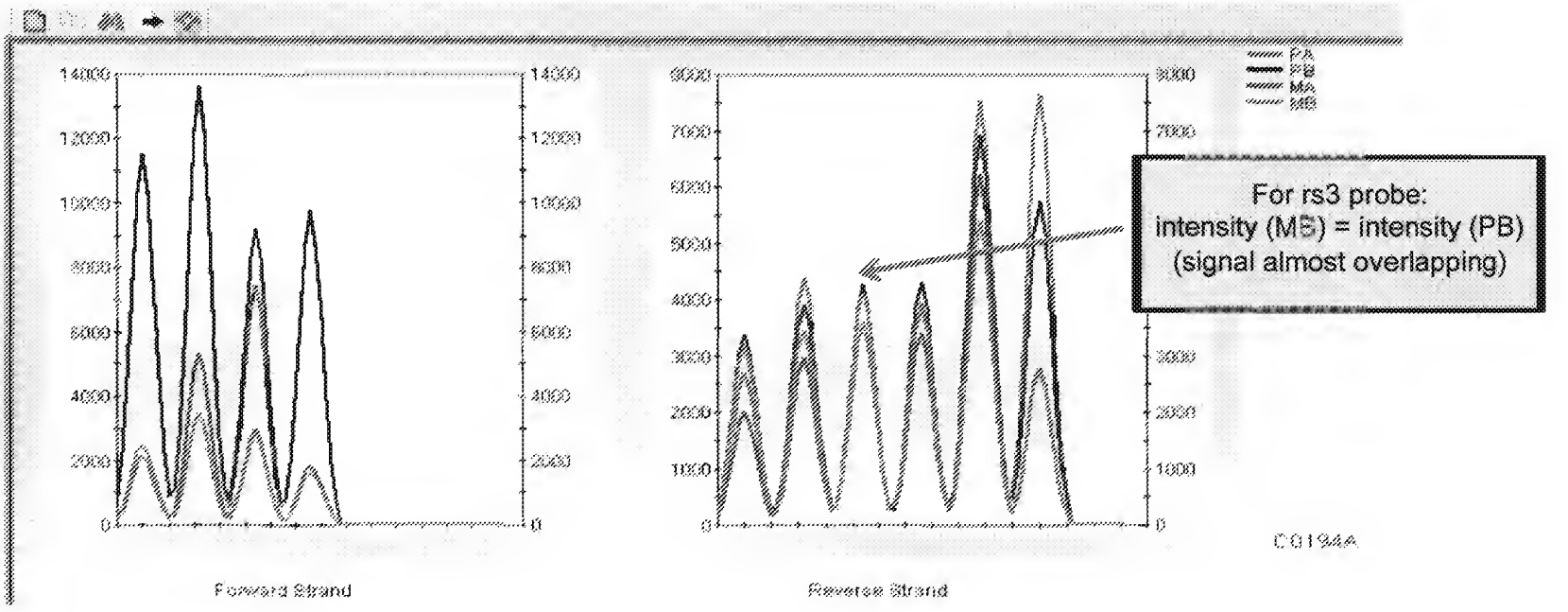
▶ No. of 25-mer probes used : (5 fs + 5 rs) X 4 (i.e. quartet for PA, PB, MA, MB) = 40

▶ PA = perfect match for A; PB = perfect match for B ; MA = mismatch for A; MB = mismatch for B

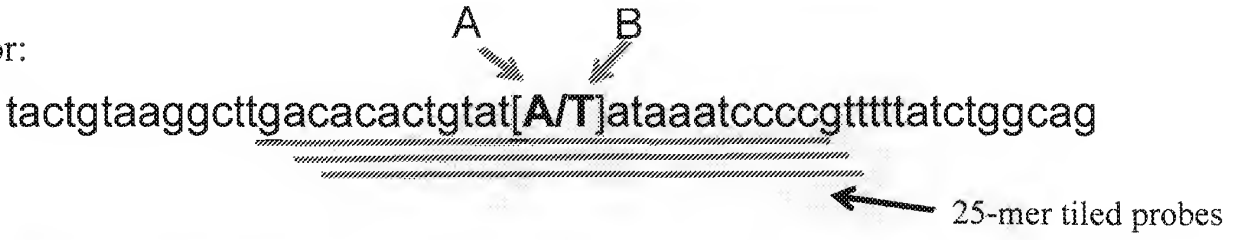
▶ GTYPE call: "BB" = **GG (homozygous guanine)**

Xba 240 SNP array analysis of CRC sample C0194A

▶ Testing the genotype for SNP_A-1646260



▶ Probesets designed for:



▶ No. of 25-mer probes used : (5 fs + 5 rs) X 4 (i.e. quartet for PA, PB, MA, MB) = 40

▶ PA = perfect match for A; PB = perfect match for B ; MA = mismatch for A; MB = mismatch for B

▶ GTYPE call: "AB" = **AT (heterozygous adenine/thymine)**

Disadvantages of Hybridization Arrays for Detecting Mutations

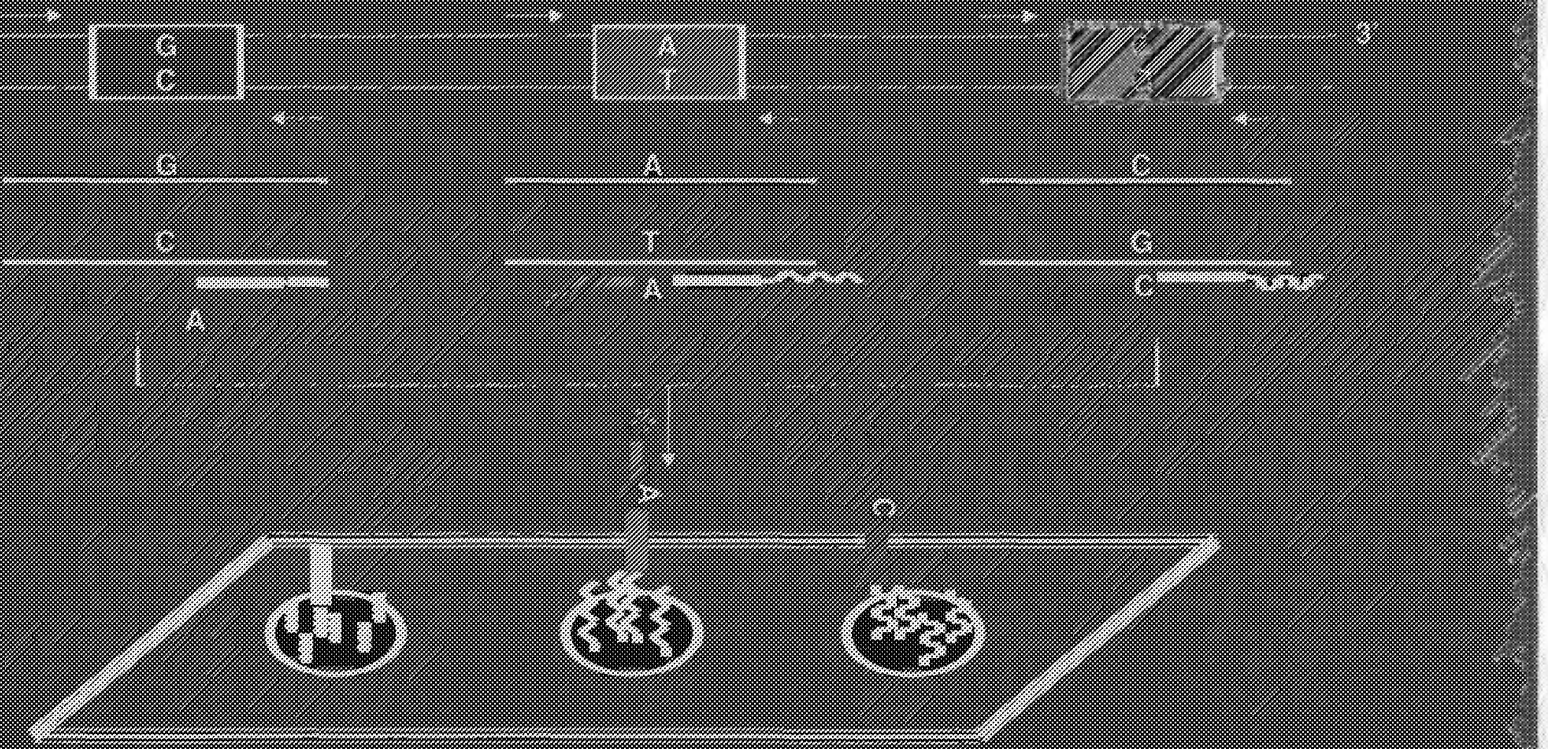
- Large variation in sequence context and probe Tm make it impossible to distinguish all mutations under uniform hybridization conditions.

- Normal DNA will sometimes hybridize to mutant probe sequence as well as the normal probe sequence, leading to false positives.

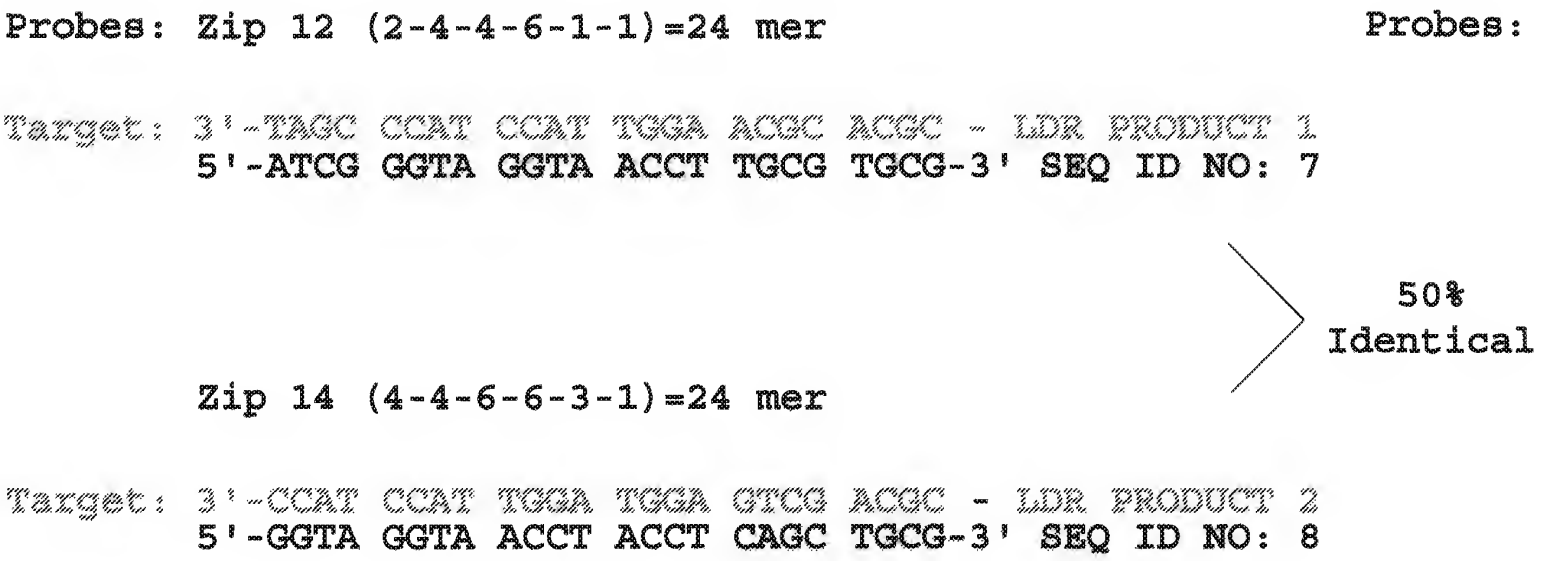
- Probes miss small deletions or insertions, leading to false negatives.

- Hybridization arrays cannot distinguish mismatch hybridization of normal DNA to mutant probe – resulting in 25%-50% signal – from match hybridization of mutant DNA, present at 25%-50% of the sample. This lack the sensitivity required to identify cancer mutations in the presence of stromal cells, leads to false negatives.

Universal Array Schematic



Identifying Mutations by Zipcode Arrays: 1



Identifying Mutations by Zipcode Arrays: 2

Probes: Zip 12 (2-4-4-6-1-1)=24 mer

Target: 3'-TAGC CCAT CCAT TGGA ACGC ACGC - LDR PRODUCT 1
5'-ATCG GGTA GGTA ACCT TGCG TGCG-3' SEQ ID NO: 7

24/24 match
Yes hybridization

Probes:
25% or more
Different

Zip 14 (4-4-6-6-3-1)=24 mer

Target: 3'-TAGC CCAT CCAT TGGA ACGC ACGC - LDR PRODUCT 1
5'-GGTA GGTA ACCT ACCT CAGC TGCG-3' SEQ ID NO: 8

12/24 match
No hybridization

Target: 3'-TAGC CCAT CCAT TGGA AC_GC ACGC - LDR PRODUCT 1
5'-GGTA GGTA ACCT ACCT CAGC TGCG-3' SEQ ID NO: 8

13/24 match
No hybridization

Grace's Family History:

1. Grandfather died at 30 (Rare Sarcoma).
2. Aunt died at 20. (Brain tumor)
3. Mother survived breast cancer at 22.

Does Grace carry a p53 mutation?

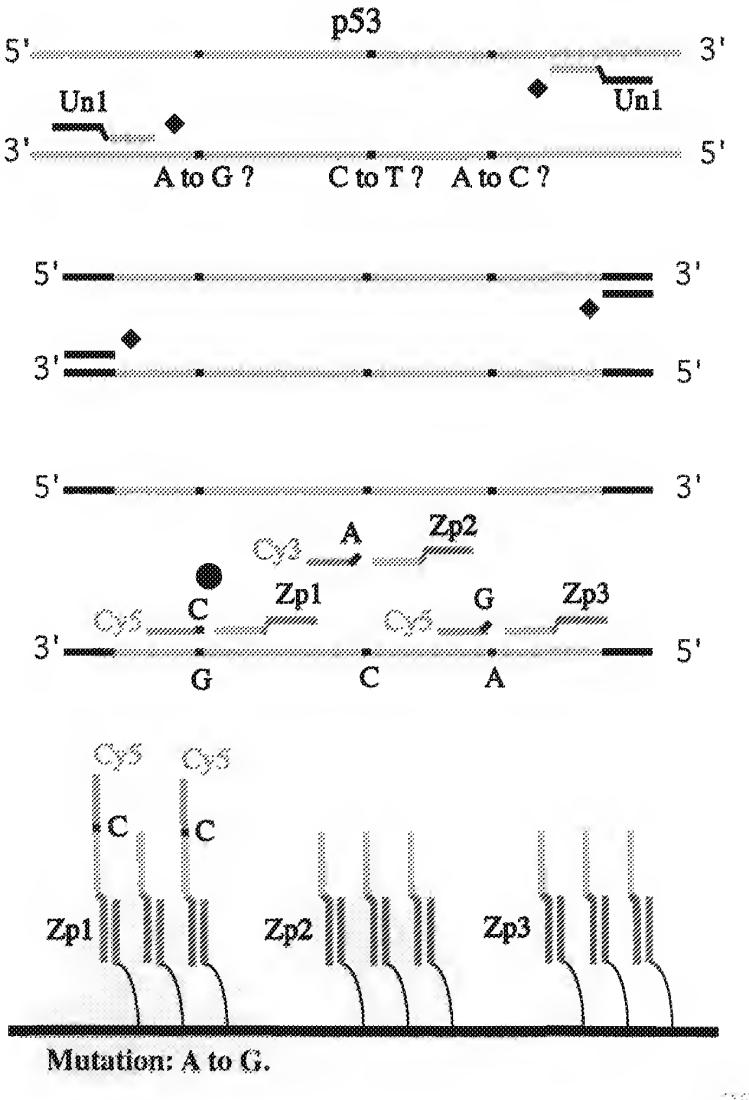
PCR / PCR / LDR / Universal Array

1. PCR amplify all p53 exons using gene-specific/universal primers and Taq polymerase. ◆

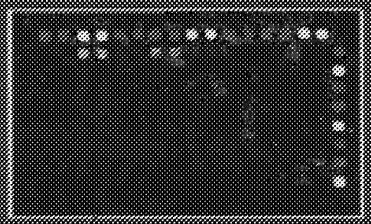
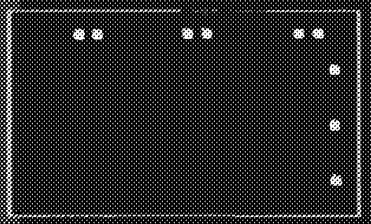
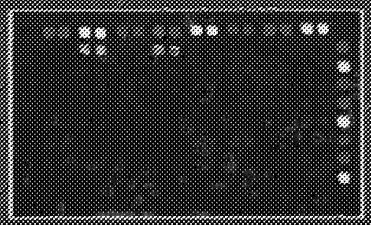
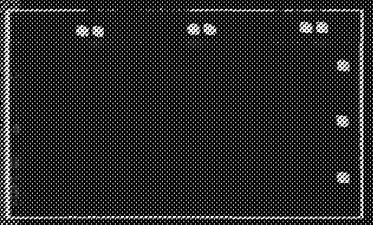
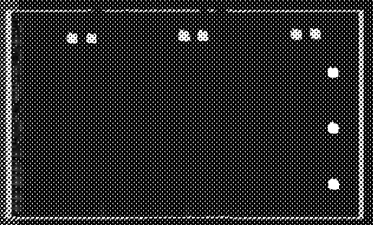
2. PCR amplify all primary products using universal primers and Taq polymerase. ◆

3. Perform LDR using mutation-specific LDR primers, common primers containing complementary zip code sequences, and thermostable ligase. ●

4. Capture fluorescent products on addressable array and score for presence of mutation.

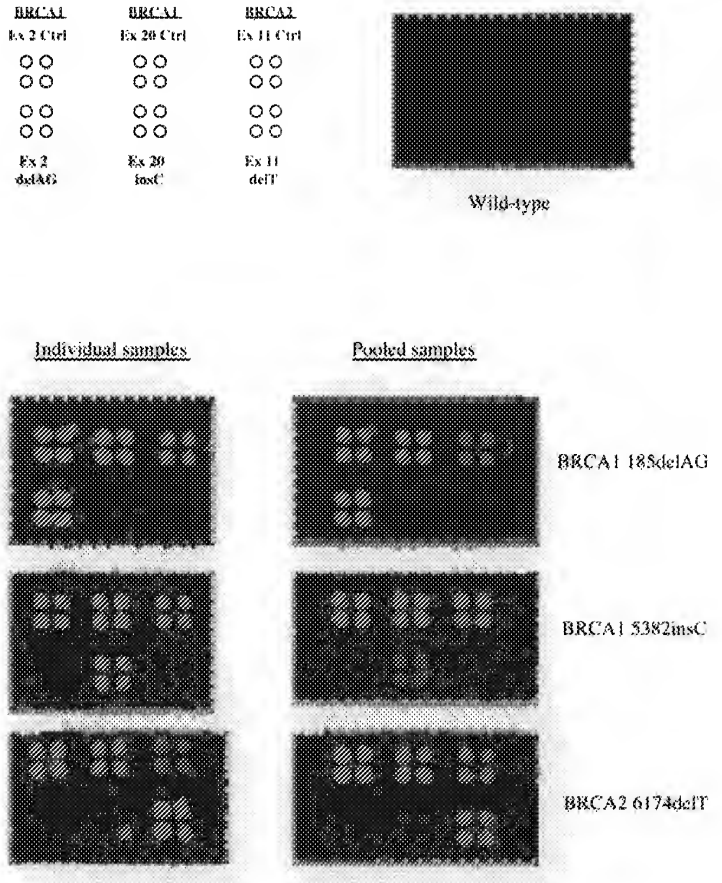


Detection of Mutations in Colon Tumor DNA using the p53 Chip

	Upper Strand	Lower Strand	Mutation
Sample 1			R273 C-T Lower strand Zip 46
Sample 2			Y220 A-G Lower strand Zip 31
Sample 3			R175 G2-A Upper strand Zip 21

Cy 3 Cy 5 FAM

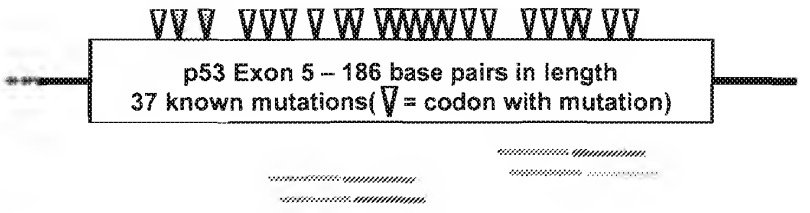
Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2



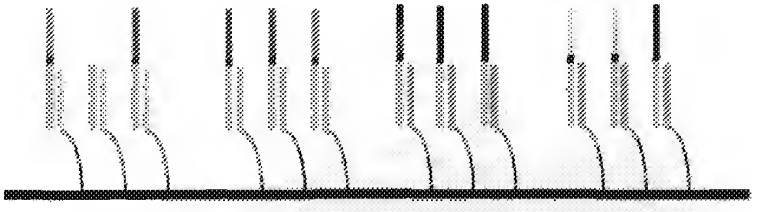
LDR detection of three specific mutations in BRCA1 and BRCA2 on an addressable universal microarray. The diagram at upper left shows the assignment of each control and mutant sequence to specific addresses on the array surface. Control signals are directed to the upper three addresses; mutant signals are assigned to the lower three. The upper right image shows signal produced by a wild-type DNA. Left panel: representative hybridizations for individual DNA samples. Right panel: representative hybridizations for each mutation using pooled samples of DNA from Ashkenazi individuals. The mutations are identified on the extreme right.

Favis R, Day JP, Gerry NP, Phelan C, Narod S, Barany F. Nat Biotechnol. 2000 May;18(5):561-4.

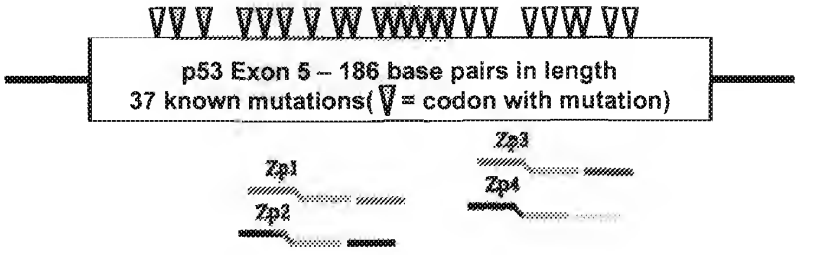
Comparison of Array Hybridization Results When Mutations are Closely Spaced



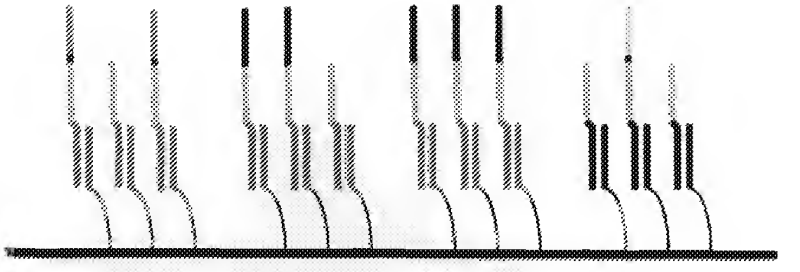
Hybridization capture probes are complementary to unlabeled primer



Results: Incorrect Signals due to Cross-Hybridization



Hybridization capture Probes are Zipcodes



Results: Only Correct Signal at Each Zipcode

Advantage of Universal Arrays

1. Provide single array design to detect all target nucleic acid differences under uniform hybridization conditions.

2. Distinguish closely spaced and overlapping mutations, including small insertions and deletions.

3. Identify low abundance mutations for early detection of cancer.

4. Identify promoter methylation silencing of tumor suppressor genes. Helps predict outcome and guide cancer treatment.

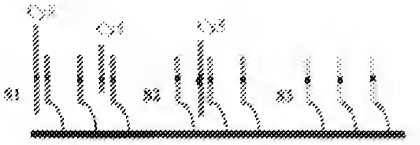
5. Identify and quantify splice site changes.

6. Quantify RNA levels for gene expression profiling.

7. Determine DNA copy level changes, loss of heterozygosity (LOH), and SNPs for genome-wide association studies. Helps predict outcome and guide cancer treatment.

Advantage of Universal Arrays over Hybridization Arrays

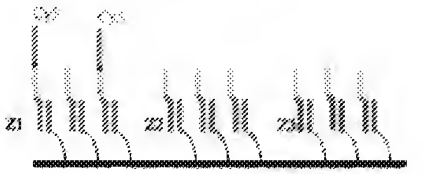
Hybridization Arrays:



• Discrimination based on match vs. mismatch hybridization; probes are very similar in sequence.

• Cross-hybridization between perfect match and mismatch probes leads to false-positives and false-negatives.

Universal Arrays:



• Capture-specific probe sequences designed to differ by 25% or more to minimize cross-hybridization.

• Composite probes or products containing a target-specific portion and a capture-specific portion allow for accurate target identification.